DOCUMENT RESUME

ED 468 487 TM 034 397

AUTHOR	Li, Yuan H.; Schafer, William D.
TITLE	Accuracy of the Generalizability-Model Standard Errors for the Percents of Examinees Reaching Standards.
PUB DATE	2002-04-00
NOTE	24p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 2-4, 2002).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	*Cutting Scores; *Error of Measurement; *Generalizability Theory; Simulation; Standards; Test Results
IDENTIFIERS	Unbalanced Designs

ABSTRACT

An empirical study of the Yen (W. Yen, 1997) analytic formula for the standard error of a percent-above-cut [SE(PAC)] was conducted. This formula was derived from variance component information gathered in the context of generalizability theory. SE(PAC)s were estimated by different methods of estimating variance components (e.g., W. Yen's balanced-sample method; Full-sample unbalanced method), and then compared with that yielded from the empirically replication-based approach. The adequacy of these methods for extending the technique to entity sizes (districts, the state as a whole) beyond the range of those used for variance component estimation (schools) was also closely examined. The data used in this simulation were from a statewide sample of students in Maryland, a state that regularly reports SE(PAC) figures. This study suggested that the full-sample unbalanced method not only produced SE(PAC)s of schools, school districts, and the whole state similar to those from Yen's balanced method, but it also resulted in less variation of SE(PAC) estimates. (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 468 487

Accuracy of the Generalizability-Model Standard Errors for the Percents of Examinees Reaching Standards

by
Yuan H. Li
Prince Georges County Public Schools, Maryland

William D. Schafer
University of Maryland at College Park

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Y. H. Li

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paper presented at the annual meeting of the National Council on Measurement in Education, April 2-4, 2002, New Orleans, LA.

BEST COPY AVAILABLE

Accuracy of the Generalizability-Model Standard Errors for the Percents of Examinees Reaching Standards

Abstract

An empirical study of the Yen (1997) analytic formula for the standard error of a percent-above-cut [SE(PAC)] was conducted. This formula was derived from variance component information gathered in the context of generalizability theory. SE(PAC)s were estimated by different methods of estimating variance components (e.g., Yen's balanced-sample method, Full-sample unbalanced method) and then compared with that yielded from the empirically replication-based approach. The adequacy of these methods for extending the technique to entity sizes (districts, the state as a whole) beyond the range of those used for variance component estimation (schools) was also closely examined.

The data used in the simulation were from a statewide sample of students in Maryland, a state that regularly reports SE(PAC) figures. This study suggested that the full-sample unbalanced method not only produced similar SE(PAC)s of schools, school districts and the whole state as the Yen's balanced method, but also resulted in less variation of SE(PAC)'s estimates.

Key Words: Standard Errors, Generalizability Theory, Variance Components, Analysis of Variance (ANOVA), Simulation

I. Introduction

For accountability purposes, states commonly develop academic content and performance standards and use an annual statewide assessment to provide an external and independent measure of how each individual student, school unit or the whole state meets them. The results are commonly reported using a statistical index (called PAC) of the percentage of students above a cutoff score. Since the PAC is a statistic, the ability to estimate its sampling error is important whenever it is used in decision making. Yen (1987) has developed an analytic method of finding the standard error of PAC (SE(PAC)), but it can be cumbersome to use and its sampling properties have not be evaluated against empirically observed variation. Accordingly, we carried out a simulation study of the Yen (1997) analytic formula. We used the Yen (1997) method as presented and with some simplifying modifications. Resampling was used to evaluate observed variation for entities of different sizes (e.g., schools, districts, the whole state) in order to compare the accuracies of the different approaches to deriving SE(PAC) estimates.

The data used in this study were from a statewide sample of students in Maryland, a state that regularly reports SE(PAC) figures. Nevertheless, the results of this study apply to any similar context in which one or more forms of a test are used to evaluate the proportion of students who are above any arbitrary scale point, whether it be determined by standard setting or by some formula to determine adequate yearly progress.

A. Background of SE(PAC) in the Maryland Statewide Assessment

Since 1991, the Maryland State Department of Education (MSDE) has implemented the annual Maryland School Performance Assessment Program (MSPAP, Maryland State Department of Education, 1998; Yen & Ferrara, 1997) for grades 3, 5, and 8 in all of its public schools. MSPAP assesses six content areas: Reading, Writing, Language Usage, Mathematics, Science and Social Studies. To cover the required breadth of learning outcomes in limited testing time, three non-parallel, but statistically linked test forms per content area are randomly assigned to students within a school.

The statistical index of PAC was calculated as the percent of students in a school who perform above a standard (e.g., at satisfactory level or better in MSPAP) in a content area. Yen (1997) developed the standard error of PAC, from variance component information gathered in the context of generalizability theory. Once the variance components are estimated, it is straightforward to generate the generalizability-based (or called formula-based) SE(PAC) values, using formulas presented in that paper (Yen, 1997).

For MSPAP, the estimation of the variance components was stratified (Yen, 1997). They were estimated separately for "small," "medium," and "large" school sizes for Grades 3 and 5, and adding "extra large" for Grade 8. For each size, the same numbers of target samples (e.g., 30, 60, 75, etc.) were randomly selected within schools. This was called a "balanced design" approach in the context of the analysis of variance (ANOVA). The balanced design method was implemented in order to control for some degree of bias that theoretically exists in the estimation of the SE(PAC) if the sizes of schools vary. However, some strange were observed time to time (e.g., in some cases the SE(PAC) for a larger school was larger than that for a smaller school).

MSDE attempted an estimation using all the data instead of stratifying. This was called an "unbalanced design" approach due to the presence of unequal sample sizes in the cells in the ANOVA. One of the goals of this study was to evaluate the effectiveness of a balanced vs. an unbalanced design in estimating the variance components for SE(PAC) estimation.

Several ANOVA methods exist (SPSS Inc., 1999) for estimating the variance components for an unbalanced data. The ANOVA Type I variance analyses chosen by MSDE resulted in relatively satisfactory results, in most cases the SE(PAC)s being identical to the balanced-design approach when rounded to the units digit. Of course, the strange anomalies were eliminated. MSDE is now using that method for its school-level reports.

Regarding the variance component estimates, ANOVA Type III method is often recommended for unbalanced data. The issue of whether Type I vs. Type III variance analyses produce more effective estimates of SE(PAC), was also investigated in this study.

There has been some pressure from the field to include SE(PAC) information for districts and for the state. Whether the generalization of the use of variance components to represent such large units is justified needs to be explored.

This provided the motivation for the present Monte-Carlo study. Generally, we repeatedly generated simulees' test responses grouped in ways that are like those in the schools in the state. The PAC value for each school, district or the whole state was calculated for each replication. The replication-based SE(PAC) was obtained by computing the standard deviation across the replicated PACs. This empirical SE(PAC) estimate generally becomes stable and accurate as the number of replication increases and served in this study as a comparison with the SE(PAC) derived from the different uses of generalizability theory.

B. Research Questions

There were four basic questions examined in this study. They are listed below:

(1). How well (in terms of accuracy and stability) does Yen's balanced-sample method estimate the actual SEs of PACs of schools?

(2). Can we substitute variance components from the full sample (unbalanced design) and obtain SE(PAC) estimates that compare well with the Yen method's estimates?

(3). Do the variance analyses from Type I ANOVA result in appreciably different SE(PAC) estimates than from the ANOVA Type III method?

(4). How well does the formula in Yen's paper, using the various variance components, estimate the actual SE(PAC)s of larger units (districts and the state)?

II. Overview of Statistical Procedures for Computing the SE(PAC)

A. Formula Used for Computing the SE(PAC)

An ANOVA model with Test Forms random and Schools fixed was recommended by (Yen, 1997) was chosen by MSDE for estimating the variance components that are then used for calculating the values of SE(PAC)s of schools. This model assumes that pupils are sampled from an infinite student population. According to this particular model, the SE(PAC) is calculated by the formula (Yen, 1997):

$$SE^2(PAC) = \frac{\sigma_f^2}{F} + \frac{\sigma_{sf}^2}{F} + \frac{\sigma_w^2}{n_{st}} \qquad (1)$$

where,

σ_f^2 = Variance of test forms,

σ_{sf}^2 = Variance of the interaction between schools and test forms,

σ_w^2 = Variance of the error term,

F = Number of test forms (3 for MSPAP)

n_{st} = Number of students per school.

The procedure of Varinace Components Analyis in SPSS (SPSS Inc., 1999) was used for computing variance components of Schools, Forms, along with the interaction of both variables.

B. Variance Components Analysis

As indicated in Equation 1, the generalizability-based SE(PAC) is computed based on the estimates of a variance component analysis. Several methods (e.g., Brennan, 2001; SPSS Inc., 1999) exist to estimate the variance components. The ANOVA approaches used in this study were either the Type I or Type III sums of squares for each effect. Type I sums of squares (Littell, Freund & Spector, 1998 ; SPSS Inc., 1999) result from the hierarchical decomposition ANOVA method, in which the sum of squares for each of the factors in the model is adjusted for only the factor(s) that follow(s) it in the design. In contrast, Type III sum of squares for each factor is estimated taking the other factors (including the interaction) into account. This is one reason why it is often recommended for unbalanced designs. In an ANOVA design with no missing cells, the Type III method is equivalent to the Yate's (1934) weighted squares of means method, described in Searle (1971).

The magnitude of variance of the forms effect computed from the Type I method is expected to be relatively larger than that computed from the Type III method because the sum of squares under the Type I method for the form effect includes shared variance with interaction, but the Type III method does not. The issue of whether this difference has any practical effect on computing the SE(PAC) was one of the questions explored in this study.

It has been noted that the ANOVA method for estimating variance components may produce negative values. Some possible reasons (see, Brennan, 2001; Shavelson & Webb, 1991) for their occurrence are: model misspecification, sampling error, or very small (or zero) true value of the variance.

Other approaches (e.g., Maximum likelihood, ML) to estimate the variance components have been introduced (e.g., Brennan, 2001; SPSS Advanced Models 10.0, SPSS Inc., 1999). This study did not evaluate these approaches.

III. Methodology

A. Test Data Generation and Ability Parameter Calibration

The observed student abilities on Social Studies for Grade 5 in the 1999 school year in the Maryland statewide testing were used for simulees' ability parameters. For each student's score record, test form, school and district identifiers were provided by MSDE, along with the test item parameters on the three Social Studies test forms for Grade 5 in the 1999 school year.

Student item data vectors were generated by the program RESGEN2.1 (Muraki, 1997) given ability and item parameters. Simulees' abilities were estimated from the vectors by the

PARSCALE calibration program (Muraki & Bock, 1996) with the test item parameters, provided by the MSDE, taken as fixed. Fixing item parameters was done in order to ensure the metric of simulees' abilities were identical across replications.

B. Sampling Procedure

1. Actual Sample Sizes From Schools for the Unbalanced ANOVA Design

Bootstrapping was used to resample examinees because students are a random factor in the model (Yen, 1997). More specifically, the sample size at each school was the number of examinees at the school, but for each replication, each simulee was sampled from the school's examinees with replacement. Following that, because forms are also random in the model, each simulee sampled was assigned to one of the three forms with equal probabilities, regardless of the form the student actually took. The total number of examinees was 62,725. Eleven schools with sample sizes less than 30 were removed and the 199 examinees from these 11 schools were excluded in estimating the variance components.

2. Target Sample Sizes for the Balanced ANOVA Design

In order to implement the ANOVA balanced-design procedure, first schools were sorted into three size categories (small, medium and large) as presented in Table 1 (see Yen, 1997). The target sample sizes were 30, 60 and 75. Second, for each size-category school, simulees were randomly sampled within schools to reach each target size.

Table 1.Numbers of Students and Schools Used for Estimating Variance Components

Groups		Range of n_{st}	n_{sch} or n_{dis}	n_{st}	Number of Schools and Simulees Used in Balanced ANOVAs in the First Replication Sampling		
					Target n_{st}	n_{sch} used in analyses	n_{st} used in Analysis
School Sizes	Small	$n_{st} < 50$	74	2836	30	42	1260
	Medium	$50 \leq n_{st} \leq 75$	151	9376	60	33	1980
	Large	$n_{st} > 75$	327	50513	75	271	20325
School Districts		$200 \leq n_{st} \leq 9638$	24	62725	N/A	N/A	N/A
State			1	62725	N/A	N/A	N/A

Notes:
n_{st}: Number of Students per School or District,
n_{sch}: Number of Schools,
n_{dis}: Number of School Districts

C. Replication Procedures

Using the procedure described above, student response vectors were generated and scored to yield scale scores that were compared with the state's cut point for "satisfactory" performance. The predicted PAC (PPAC) for each school, district and the state were computed as that unit's percent above the "satisfactory" cut point.

This process was repeated 100 times and the replication-based $SE(PAC)_{rep}$ statistic for each school, district and the state was calculated by:

$$SE(PAC)_{rep} = \sqrt{\frac{\sum_{i=1}^{r}\left(PPAC_i - Mean(PAC_i)\right)^2}{r}} \qquad (2)$$

where

$$Mean(PAC) = \frac{\sum_{i=1}^{r}(PPAC_i)}{r} \qquad (3)$$

r represents the number of replications.

For each replication, the following formula-based SE(PAC)s for each school, district and the whole state were calculated:

The generalizability-based $SE(PAC)_{gen_1}$ using the variance components obtained from ANOVA Type I unbalanced-design variance analyses.

The generalizability-based $SE(PAC)_{gen_3}$ using the variance components obtained from ANOVA Type III unbalanced-design variance analyses.

The generalizability-based $SE(PAC)_{gen_b}$ using the variance components obtained from ANOVA balanced-design. As indicated previously, the estimation of the variance components were estimated separately for "small," "medium," and "large" school sizes (variance components obtained from the large-sizes schools were also used to calculate the SE(PAC) of the district and the state).

D. Data Analysis & Evaluation

1. Evaluating the Accuracy of the Generalizability-based $SE(PAC)_{gen}$ Estimates

From Equation 1, the generalizability-based SE(PAC) is the function of variance components for a given ANOVA model and a school's sample size. The value of each of the variance components differed across simulations. The expected value of each variance

component was computed as the average variance estimate across replications. The expected variance components estimated by the different methods are presented in Table 2. When the negative expected value of variance occurred and its value is very close to zero, the true value of this variance was assumed to be zero. As seen in Table 2, four of the five expected variance for the interaction effect, Form x School, are minimal and negative so that they were substituted with zero while computing the overall SE(PAC) estimates. Setting the variance to be zero in Equation 1 is equivalent to dropping the corresponding interaction effect (between Test Forms and Schools) from the model. When the expected values of variance components were used in Equation 1, the generalizability-based SE(PAC) values were obtained and compared with those obtained from the replication-based SE(PAC).

Table 2. Descriptive Statistics of Variance Components across 100 Replication Estimates

Method	Variance	Mean	SD	Minimum	Maximum
ANOVA Type I	Form	0.00005450	0.00003657	-0.00000139	0.00013902
	Form x School	-0.00028941	0.00027730	-0.00117274	0.00039687
	Error	0.22436491	0.00061151	0.22263079	0.22584357
ANOVA Type III	Form	0.00004799	0.00004339	-0.00001352	0.00017555
	Form x School	-0.00028941	0.00027730	-0.00117274	0.00039687
	Error	0.22436491	0.00061151	0.22263079	0.22584357
Balanced Small Schools	Form	0.00021531	0.00062883	-0.00057724	0.00248758
	Form x School	0.00013171	0.00323155	-0.00797354	0.00867118
	Error	0.20363825	0.00694217	0.18631240	0.22020202
Balanced Medium Schools	Form	0.00007344	0.00034923	-0.00041429	0.00143756
	Form x School	-0.00017760	0.00172282	-0.00394693	0.00344424
	Error	0.20410217	0.00777953	0.18730811	0.22430673
Balanced Large Schools	Form	0.00005235	0.00006936	-0.00003170	0.00029472
	Form x School	-0.00003487	0.00056453	-0.00164051	0.00126702
	Error	0.22506318	0.00130486	0.22188406	0.22809996

2. Evaluating the Variation of the Generalizability-based SE(PAC) Estimates

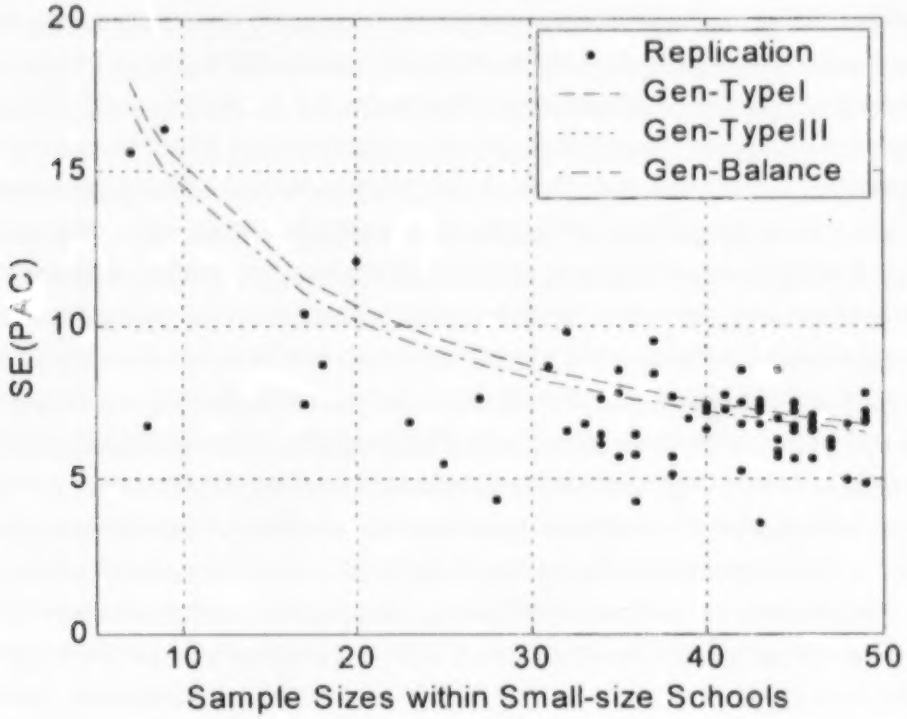
For each of the three generalizability-based approaches, the 100 generalizability-based SE(PAC)s for each school, district and the whole state were calculated as indicated in the section on procedures. When a negative variance estimate occurred, this variance estimate was set to zero. The standard deviation or the range (between maximum and minimum) of the 100 SE(PAC)s was computed for each school, school district and the state. The plot of the 95% confidence intervals of SE(PAC) (SE(PAC) ± 1.96 times the standard deviation of SE(PAC)) for all schools or school districts against their sorted sample sizes was graphed. Similarly, The plot of ranges for all schools or school districts against their sorted sample sizes was also displayed.

IV. Results and Discussion

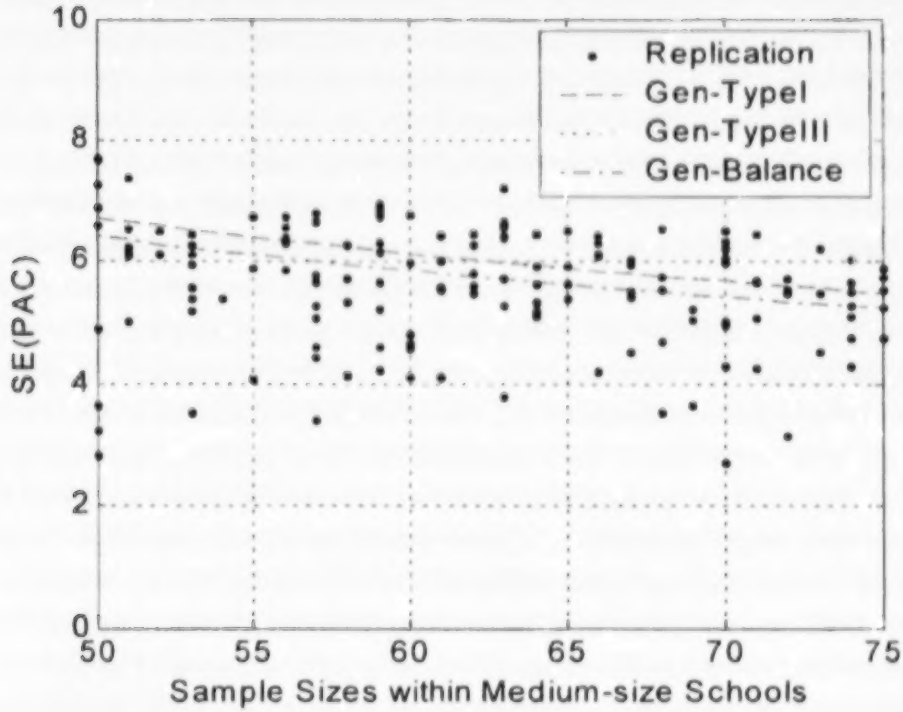
A. Issues Associated with SE(PAC) Estimates

1. Comparisons in SE(PAC) Estimates among Replication- and the Three Formula-based Methods

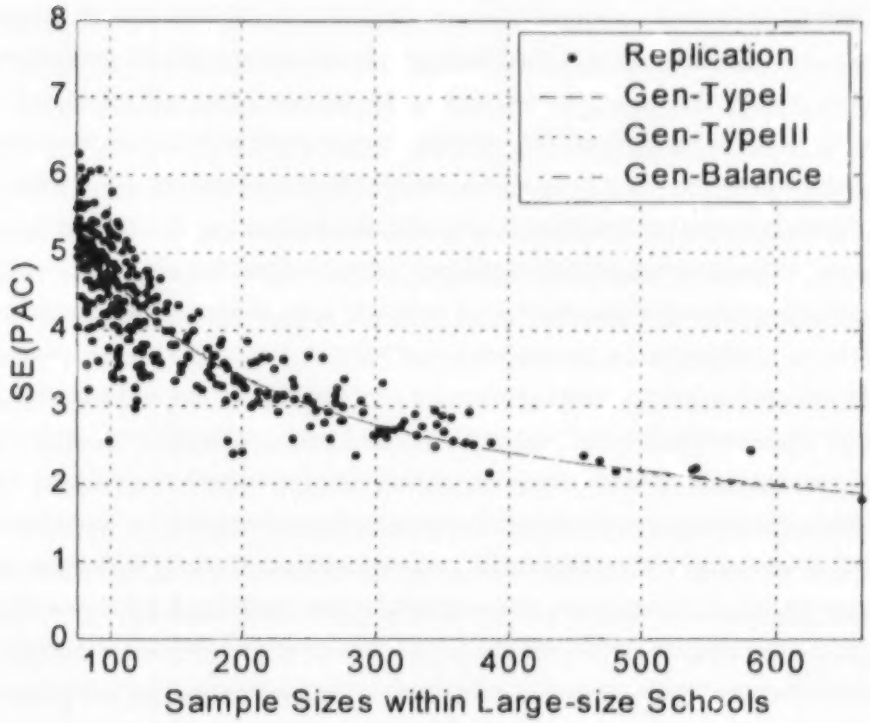
Figure 1 displays plots of SE(PAC)s of all schools (or districts) against their sample sizes (from smallest to largest) for the replication-based and three formula-based approaches. The average SE(PAC)s across small-size, medium-size, and large-sizes schools, as well as districts and the whole state are presented in Table 3.



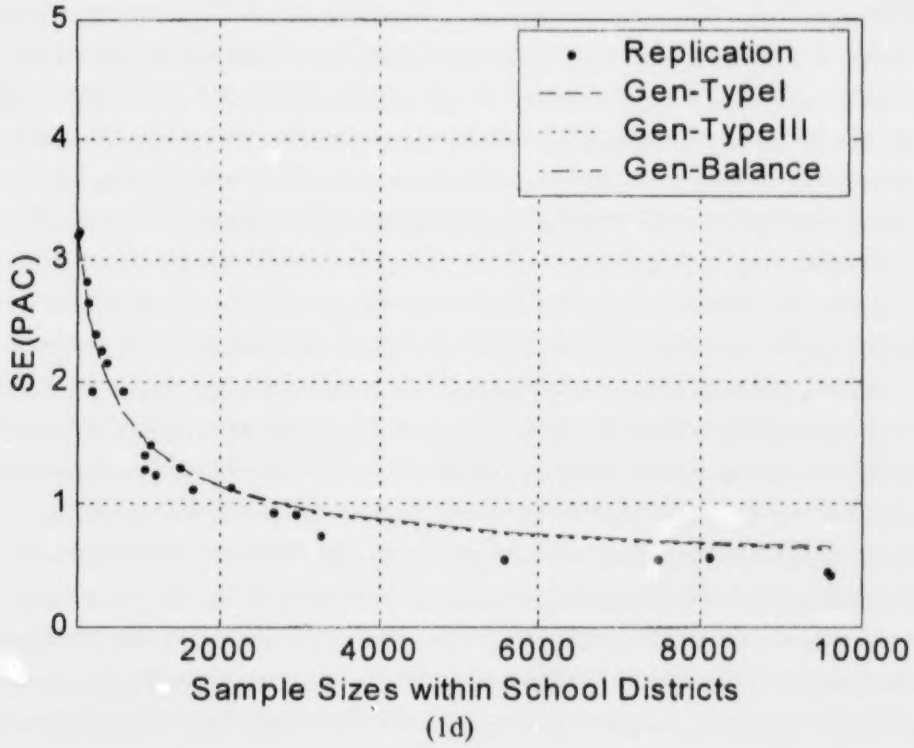
(1a)


Replication
Gen-TypeI
Gen-TypeIII
Gen-Balance
SE(PAC)
Sample Sizes within Medium-size Schools

(1b)


Replication
Gen-TypeI
Gen-TypeIII
Gen-Balance
SE(PAC)
Sample Sizes within Large-size Schools

(1c)



(1d)

Figure 1. A Plot of SE(PAC) of Schools against their Sample Sizes for the Replication-based and the Formula-based Methods

Figure 1a for Small-size Schools, Figure 1b for Medium-size Schools, Figure 1c for Large-size Schools and Figure 1d for School Districts.

Table 3. Average SE(PAC)s across Schools (Small, Medium, Large or All) or School Districts.

Types		Replication	Gen-TypeI	Gen-TypeIII	Gen-Balance
School					
Sizes	Small	7.1805	8.0734	8.0720	7.7582
	Medium	5.6262	6.0580	6.0562	5.7849
	Large	4.0776	4.2150	4.2122	4.2206
	All Schools	4.9172	5.2364	5.2341	5.1227
School Districts		1.5262	1.6000	1.5912	1.5994
State		0.1151	0.4663	0.4424	0.4587

Table 3 and the plots in Figure 1 show that ANOVA Type I vs. Type III made little difference across all sizes of schools, as well as districts and the state. This result is consistent with the finding that there were almost no interaction effects in our data (see Table

2). If substantive interaction effects occur, ANOVA Type I and Type III will produce different variance components that will likely affect the values of the SE(PAC) estimates.

Figures 1a and 1b show that the ANOVA balanced procedure was slightly different from the unbalanced procedures for small-size and medium-size schools. For the large–size schools and districts, both methods produced undistinguishable results.

On the whole, all the formula-based methods seem to be slight overestimates of the SE(PAC) compared with the replication-based SE(PAC). The formula seems to generalize quite well to large (district-level) sizes even to the whole state (refer to Table 3).

2. Modeling the Replication-based SE(PAC) Estimates

The replication-based SE(PAC) estimates could vary slightly for schools of the same size. Thus, the observed dots in Figure 1 for the replication-based SE(PAC)s did not form a smooth curve as did the formula-based SE(PAC)s. We attempted empirically to find a nonlinear model among ten standard curve-fit modes (see SPSS Inc, 1999, p. 237) in order to fit these points. A power model fit the replication-based SE(PAC) estimates best when the sample sizes of schools, districts and the state were used as independent variable:

$$\text{Model-fit SE(PAC)} = b_0(n^{b1}) \qquad (4)$$

where

b_0 and b_1 are the power model parameter estimates,

n is a school (or district or state) sample size.

When fitting the Power model to those empirically replication-based SE(PAC) estimates, the b_0 and b_1 were estimated as .3763 and -.4264. The plot of the model-fit SE(PAC)s, as well as their corresponding replication-based SE(PAC)s, against their sample sizes (schools and districts) was graphed in Figure 2. It is noted that although the state sample size was also included into the power model, the result for the state was not graphed in Figure 2 in order to retain space for displaying results for the schools and districts. Figure 2 demonstrate that the model-fit SE(PAC) estimates fit their corresponding replication-based SE(PAC)s well across sample sizes.

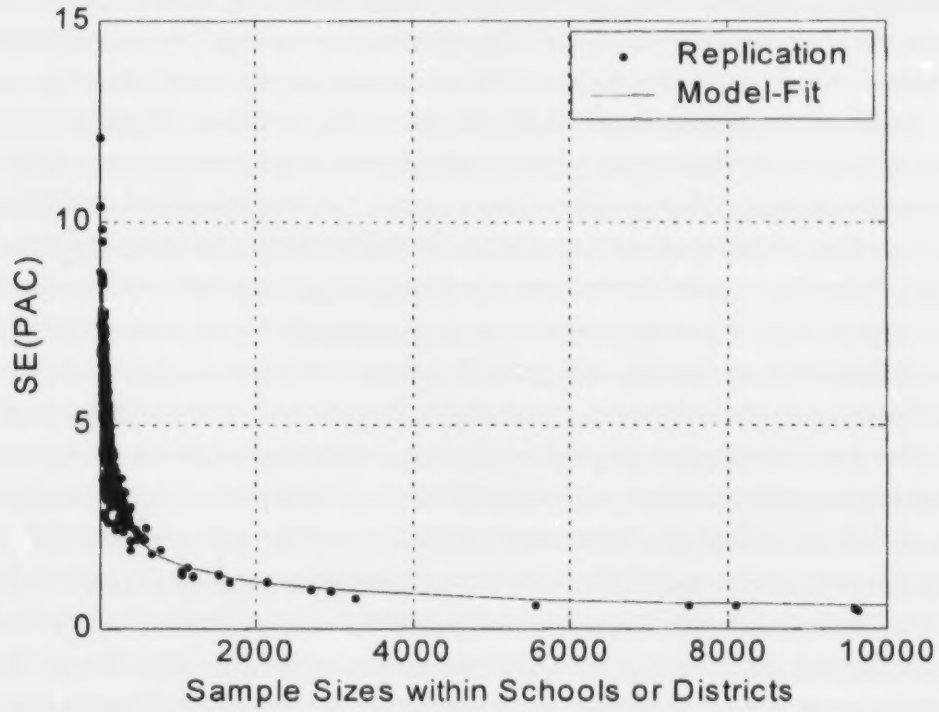


Figure 2. A Plot of SE(PAC) of Schools or Districts against their Sample Sizes for the Replication-based and the Model-fit Methods

The model-fit method also provides variation information for the model-fit SE(PAC) estimates. Figure 3 is an attempt to utilize this information. Figure 3 plots the X-axis from the lowest school size up to the largest district size. SE(PAC) is on the Y axis. Six curves are plotted: (1) the model-fit SE(PAC)s, (2) the model-fit SE(PAC)s plus one standard deviation of the model-fit SE(PAC)s, (3) the model-fit SE(PAC)s minus one standard deviation of the model-fit SE(PAC)s, (4) the ANOVA Type I-based SE(PAC)s, (5) the ANOVA Type III-based SE(PAC)s, and (6) the ANOVA balance-based SE(PAC)s. Figure 3 shows that the three formula-based SE(PAC) estimates across sample sizes are located within the range between above one standard deviation and below one standard deviation the model-fit SE(PAC) estimates.

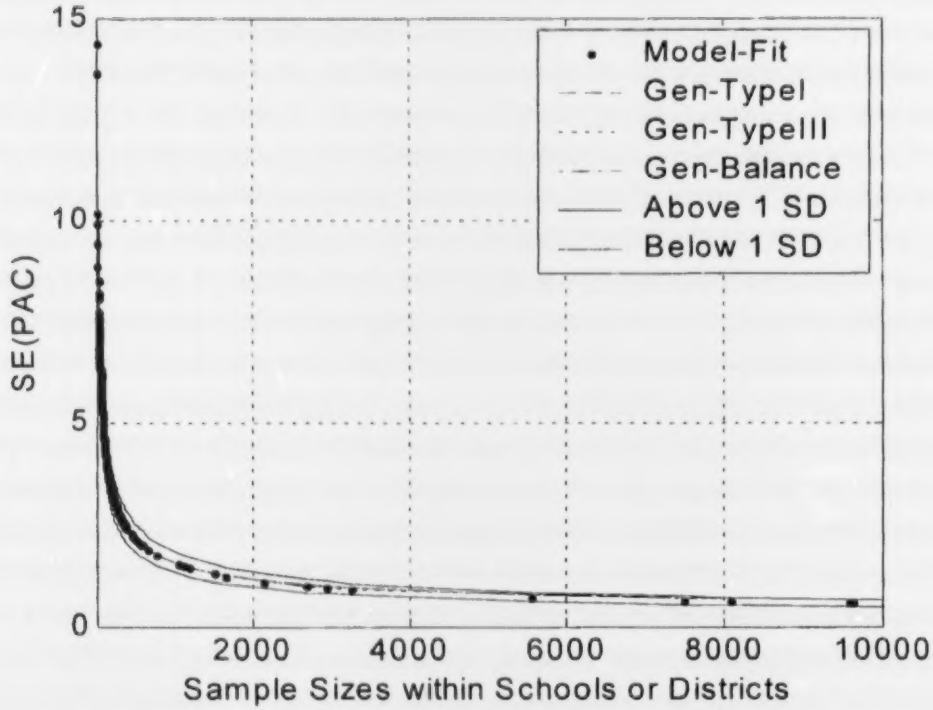


Figure 3. A Plot of SE(PAC) of Schools or Districts against their Sample Sizes for the Model-fit, the Formula-based Methods, as well as the ± 1 SD of the Model-Fit SE(PAC).

3. The Impact of a School's PAC value on Its SE(PAC)

One might wonder whether the school's PAC value has any effect on its SE(PAC). Figure 4 is a 3D graph showing relationships among school's replication-based SE(PAC), the school's sample size (N) and School's PAC. Figure 4 does not seem to show any patterns. It does not appear that the larger the PAC, the larger (or smaller) the SE(PAC). In other words, a school's PAC may not be a promising predictor for its variation of the PAC estimate.

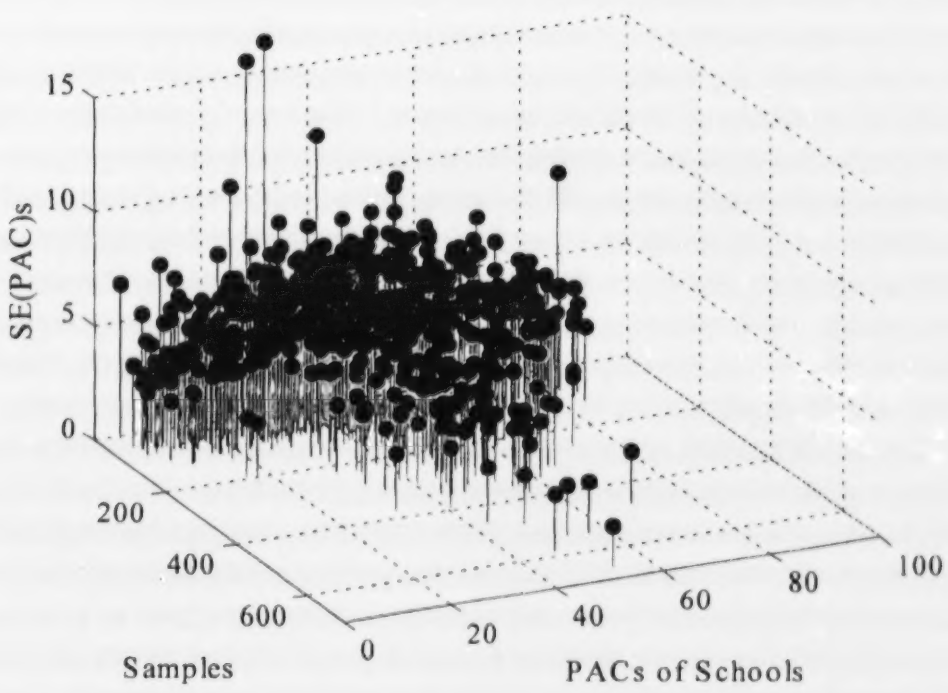


Figure 4. A 3-D Graph to Demonstrate the Inter-Associations Among SE(PAC)s of Schools, PAC of Schools and Sample Sizes of Schools.

B. Comparisons in Variation of the Three Formula-based SE(PAC) Estimates

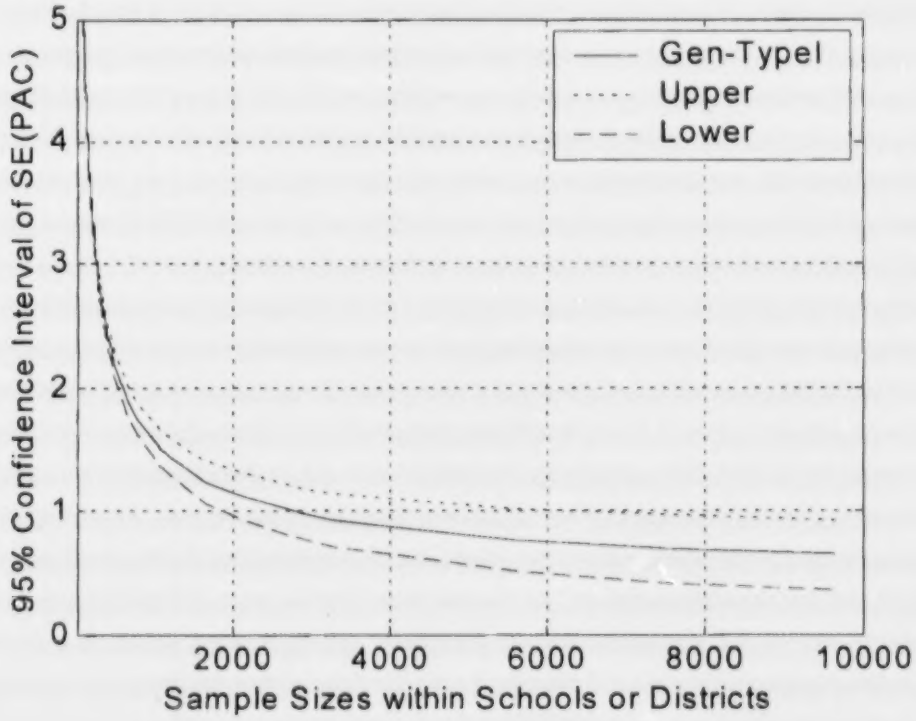
1.Plots for 95% Confidence Intervals

Figure 5 displays plots of 95% confidence intervals of the formula-based SE(PAC)s for all schools (or districts) against their sample sizes (from smallest to largest). Figures 5a, 5b and 5c show the ANOVA Type I, TYPE III and Balance-Design approaches, respectively. A comparison between Figures 5a and 5b indicates that, in general, Type I vs. Type III sums of squares make no practical difference in the variations of SE(PAC) estimates across school (or district) sample sizes. The 95% confidence intervals of the formula-based SE(PAC) estimates from the ANOVA unbalanced methods seem relatively small, but that is not the case for the ANOVA balanced method (see Figure 5c). As noted earlier, the balanced design makes use of a much smaller proportion of examinees (refer to Table 1) for estimating the variance components, which is likely the major reason for the differences in stability of SE(PAC) estimates.

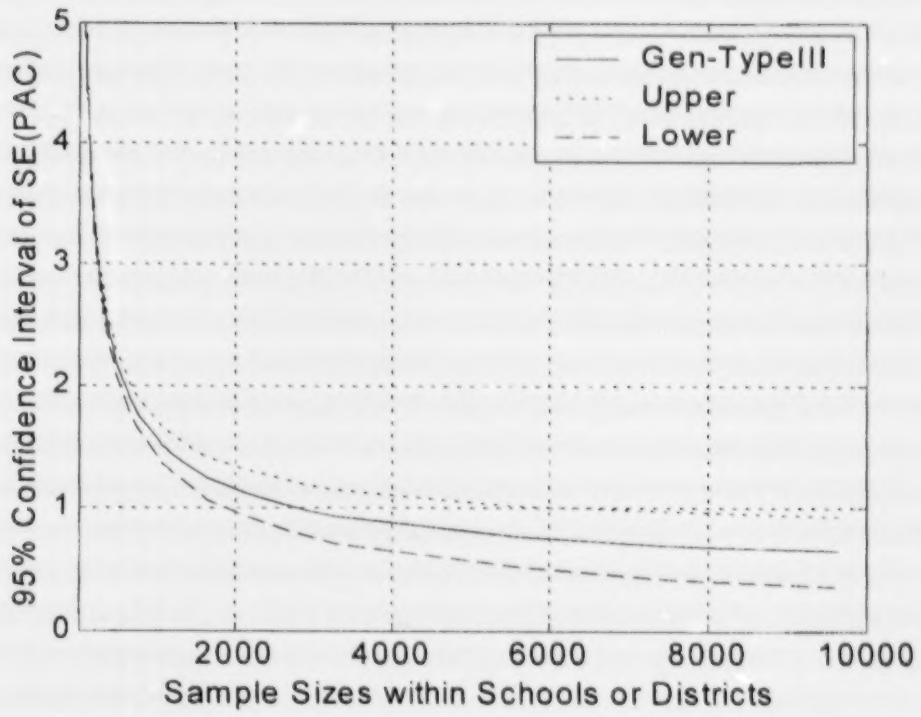
The 95% confidence interval of SE(PAC) in Figure 5 was computed solely on a school-by-school basis. In other words, for each school we used the standard deviation of its

own 100 SE(PAC)s to compute its 95% confidence interval of SE(PAC) estimates. Figure 5 shows that a school with a small sample size produces relatively homogeneous formula-based SE(PAC) estimates across 100 replications. In contrast, a school with a larger sample size produces more varied formula-based SE(PAC) estimates across replications. The districts have even larger sample sizes so that the values of their SE(PAC) are relatively smaller, but the values of variances of their formula-based SE(PAC) estimates are relatively larger.

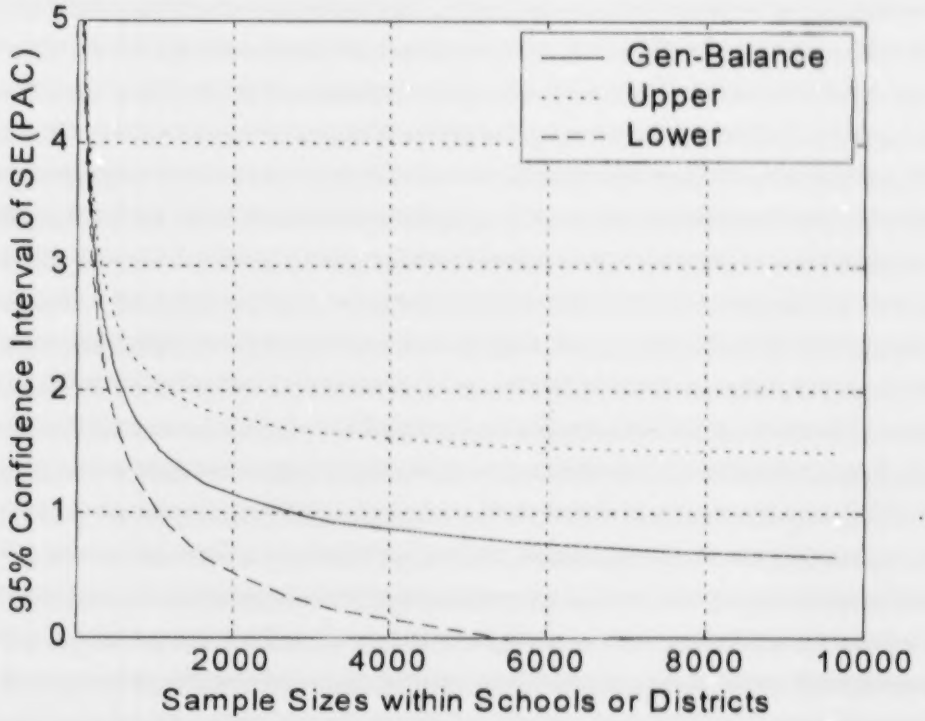
In short, the variability of the formula-based SE(PAC) is increasing with the sample size. As seen in Equation 1, the residual term in the ANOVA model is better estimated because of with more degrees of freedom than any of the other terms. But with larger samples sizes, it contributes proportionally less to the SE(PAC) estimate. So the less stable terms have grater weight for larger sample sizes.



(5a)


Gen-TypeIII
Upper
Lower
95% Confidence Interval of SE(PAC)
0
1
2
3
4
5
2000
4000
6000
8000
10000
Sample Sizes within Schools or Districts

(5b)


Gen-Balance
Upper
Lower
95% Confidence Interval of SE(PAC)
0
1
2
3
4
5
2000
4000
6000
8000
10000
Sample Sizes within Schools or Districts

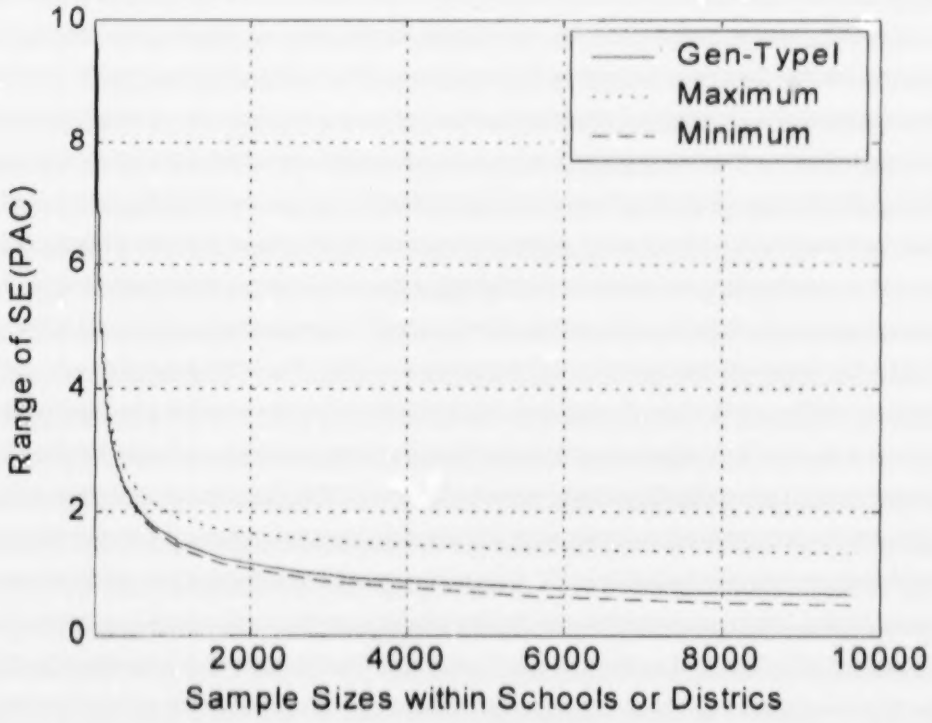
(5c)

Figure 5. A Plot of 95% Confidence Interval of SE(PAC) Estimates for Schools or Districts against their Sample Sizes

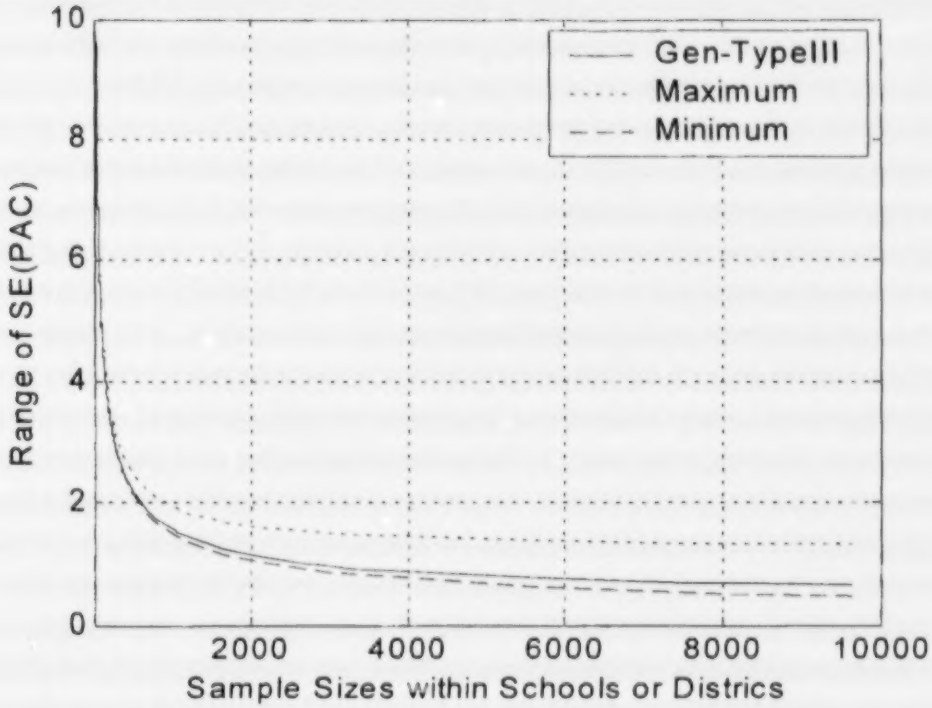
Figure 5a for ANOVA Type I, Figure 5b for ANOVA Type III, and Figure 5c for ANOVA Balanced

2.Plots for the Range Between Maximum and Minimum SE(PAC)s

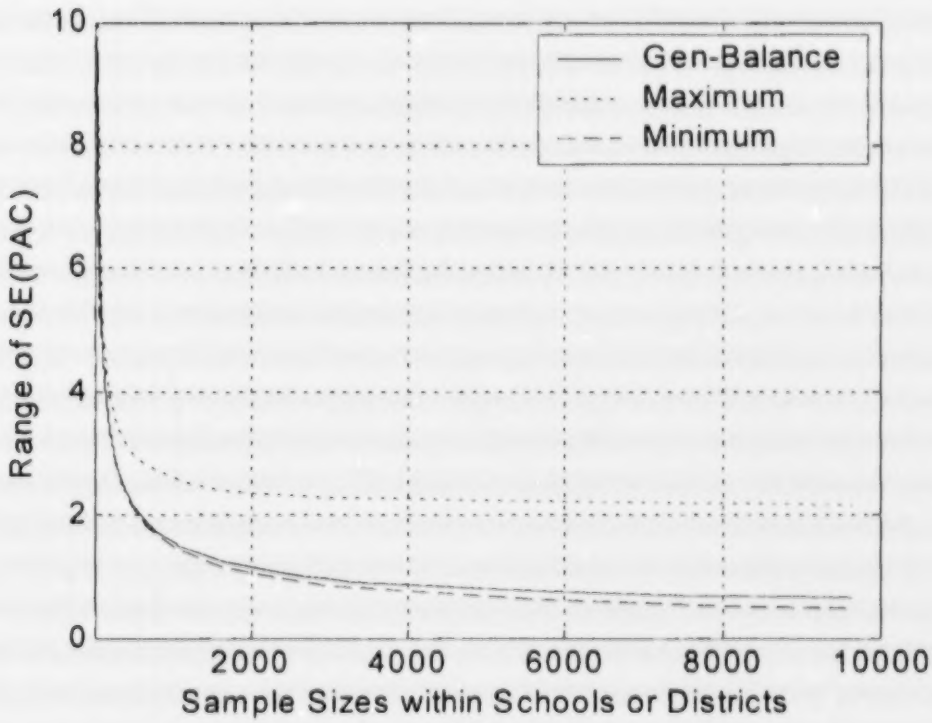
Figure 6 are the plots of the range between maximum and minimum SE(PAC)s of all schools (or districts) against their sample sizes for the three generalizability-based approaches. Figure 6 also shows that, in general, Type I vs. Type III sums of squares make no difference in the ranges of SE(PAC) across sample sizes. The ranges from the ANOVA unbalanced method across schools were relatively smaller than the ANOVA balanced method (see Figure 6c).



(6a)



(6b)



(6c)

Figure 6. A Plot of Range between Maximum and Minimum of SE(PAC) Estimates for Schools or Districts against their Sample Sizes

V. Conclusions

Yen's balanced-sample method appears to produce relatively consistent results compared with those from the replication-based method. However, her approach appears to produce less stable SE(PAC) estimates than the unbalanced full-sample methods, likely due to the fact that this design can only use part (here, about 37.57%) of the examinee's test scores for estimating the variance components.

This study suggests that the full-sample unbalanced method not only produce similar SE(PAC)s of schools, school districts and the whole state as the Yen's balanced method, but also yield less variation of SE(PAC)'s estimates. Utilizing all examinee's test scores to estimate variance components appears to be a key factor that makes the unbalanced design yield relatively reliable estimate of SE(PAC).

The ANOVA Type I and Type III sums of squares resulted in almost identical SE(PAC) estimates under this study's scenario. This conclusion can only apply to the model and test data similar to this study. Because the interaction is not taken into account in estimating the variances of main effects using the ANOVA Type I approach, it is problematic if the interaction is significant. Therefore, ANOVA Type III method should be preferred in general.

Because the SE(PAC)s yielded from the formula-based approaches are close to those produced from the replication-based approach, the formula-based method (especially for the unbalanced design) can be used practically in estimating the SE(PAC) of schools. This conclusion can generalize to the large sample sizes such as school districts and the entire state. However, as a sample size of a unit (school, district) increases, the SE(PAC) of the unit becomes gradually smaller, but the variation of formula-based SE(PAC) estimates for the unit becomes larger. This finding reminds us to be cautious when reporting the SE(PAC) for a large unit.

Figure Headings

Figure 1. A Plot of SE(PAC) of Schools against their Sample Sizes for the Replication-based and the Formula-based Methods

Figure 1a for Small-size Schools, Figure 1b for Medium-size Schools, Figure 1c for Large-size Schools and Figure 1d for School Districts.

Figure 2. A Plot of SE(PAC) of Schools or Districts against their Sample Sizes for the Replication-based and the Model-fit Methods

Figure 3. A Plot of SE(PAC) of Schools or Districts against their Sample Sizes for the Model-fit, the Formula-based Methods, as well as the ± 1 SD of the Model-Fit SE(PAC).

Figure 4. A 3-D Graph to Demonstrate the Inter-Associations Among SE(PAC)s of Schools, PAC of Schools and Sample Sizes of Schools.

Figure 5. A Plot of 95% Confidence Interval of SE(PAC) Estimates for Schools or Districts against their Sample Sizes

Figure 5a for ANOVA Type I, Figure 5b for ANOVA Type III, and Figure 5c for ANOVA Balanced

Figure 6. A Plot of Range between Maximum and Minimum of SE(PAC) Estimates for Schools or Districts against their Sample Sizes

Figure 6a for ANOVA Type I, Figure 6b for ANOVA Type III, and Figure 6c for ANOVA Balanced

References

Brennan, R. L. (2001). Generalizability theory. New York: Springer-Verlag.

Littell, R. C., Freund, R. J. & Spector, P. C. (1998). SAS system for linear models, third edition. Cary, NC: SAS Institute Inc.

Muraki, E. (1997). RESGEN2.1: Item response generator. Educational Testing Service.

Muraki, E. & Bock, R. D. (1996). PARSCALE (Version 3.): IRT based test scoring and item analysis for graded open-ended exercises and performance tasks. Mooresvilk: Scientific Software.

Searle, S. R. (1971). Linear models. New York: John Wiley & Sons, Inc.

Shavelson, R.J. & Webb, N. M. (1991). Generalizability theory: A Primer. Newbury Park,CA: SAGE Publications, Inc.

SPSS Inc. (1999). SPSS Advanced models 10.0. [Computer program]. Chicago, IL: SPSS Inc.

Yates, F. (1934). The analysis of multiple classifications with unequal numbers in the different classes. Journal of the American Statistical Association, 29, 51-66.

Yen, W. M. (1997). The technical quality of performance assessments: Standard errors of percents of pupils reaching standards. Educational Measurement: Issues and practice, 16, 5-15.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

ERIC Educational Resources Information Center

TM034397

I. DOCUMENT IDENTIFICATION:

Title:	**Accuracy of the Generalizability-Model Standard Errors for the Percents of Examinees Reaching Standards**
Author(s):	**Yuan H. Li & William D. Schafer**
Corporate Source:	Publication Date: 2002

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☒

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only.

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only.

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please →

Signature: Yuan Li	Printed Name/Position/Title: Yuan H. Li; Statistical Specialist
Organization/Address: Prince George's County Public Schools Testing Office, Room, 202E Upper Marlboro, MD. 20772	Telephone: 301-952-6964 FAX: 301-952-6228 E-Mail Address: yuanhwangli@juno.com Date: 9/5/2002

(Over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfacility.org